Age and Gender Detection

Hemant Patel[1], Saurabh Mittal[1], Sahil Choudhary[1]

Dr. Vibhakar Phathak[2], Dr. Akhil Panday[3],Dr. Vishal Shrivastsva[4], Dr. Krishan Kant Lavania[5]

B.TECH. Scholar, Department of Information Technology[1]

Professor, Department of Information Technology[2-5]

Arya College of Engineering & I.T. Kukas, Jaipur, India

[1]hemantpatel272001@gmail.com, saurabhmittal89788@gmail.com, sahilchoudhary1039@gmail.com

[2]vibhakar @aryacollege.in, [3]akhil@aryacollege.in , [2]vishalshrivastava.cs@aryacollege.in,

[5]krishankantlavania@aryacollege.in

Abstract: *Facial palsy (FP) is a neurological disorder that affects the facial nerve, specifically the seventh nerve, resulting in the patient losing control of the facial muscles on one side of the face. It is an annoying condition that can occur in both children and adults, regardless of gender. Diagnosis by visual examination, based on differences in the sides of the face, can be prone to errors and inaccuracies. The detection of FP using artificial intelligence through computer vision systems has become increasingly important. Deep learning is the best solution for detecting FP in real-time with high accuracy, saving patients time, effort, and cost. Therefore, this work proposes a real-time detection system for FP, and for determining the patient's gender and age, using a Raspberry Pi device with a digital camera and a deep learning algorithm. The solution facilitates the diagnosis process for both the doctor and the patient, and it could be part of a medical assessment activity. This study used a dataset of 20,600 images, containing 19,000 normal images and 1600 FP images, to achieve an accuracy of 98%. Thus, the proposed system is a highly accurate and capable medical diagnostic tool for detecting FP.*

Keywords: facial palsy; raspberry Pi; real-time; face detection; computer vision system; artificial intelligence Estimating the age, classifying gender, Raspberry Pi 5, learning the depths and breadths of deep learning, facial feature extraction, real-time inference, understanding OpenCV, understanding transfer learning, edge computing, our understanding of embedded AI, smart surveillance tech

I. INTRODUCTION

Facial palsy is a facial nerve disease that occurs on one side of the face and leads to loss of control of voluntary muscle movement [1]. There are a range of symptoms associated with FP, including taste and hearing problems, pain around the face and in the ears, sagging eyelids, and dry eyes. FP affects one out of every 60 people [2,3]. The people most at risk of developing FP are pregnant women, diabetics, and those with a family history, with an incidence of around 10% [4]. The percentage of infection on the left side of the face is higher than on the right side, and the percentage of afflicted males is lower than that of females [5]. Statistically, 6.1 out of every 100,000 people with FP are children, with an average age of 1 to 15 years [2]. FP is most common between the ages of 30 and 45 years, and it annually affects 37.7 out of every 100,000 people in the UK [6]. The traditional method still prevails in the medical diagnosis of FP, which depends on the doctor's vision and judgment and requires the patient to spend time, effort, and money. FP is annoying and uncomfortable because of the deformation that occurs to the face and the accompanying pain and other symptoms. Therefore, an automatic system should be developed to detect FP accurately and fast.

Machine learning is a branch of artificial intelligence that makes predictions about data through model learning, allowing computers to gather insights and learn about data [7]. A convolutional neural network (CNN) performs complex pattern and image recognition tasks and is a form of artificial intelligence [8]. CNN differs from the mainstream data feature extraction methods because it can extract features using convolutional structures. Recognizing objects through computer vision is the operation that CNN was created to perform [9,10,11]. Deep learning is one of the

most advanced solutions to the problem of face detection and recognition. It solves the problems of digital face image processing applications, such as the problem of image colorization, detection, and classification [3,12,13]. Computer vision is a method of collecting and interpreting visual information by a computer, essentially teaching a computer to see things [14].

In recent years, after the shift in FP detection technology from traditional methods to automatic detection methods by means of machine learning and computer vision, several methods of FP detection appeared. For example, a study by Dong et al. [15] measured the degree of FP by identifying the difference between faces and proposed a quantitative estimation method to detect the main points by K-MEANS clustering and using the Susan edge algorithm to identify the salient points of the edges and detect facial features. Another study by Azoulay et al. [16] worked on detecting vital signs and diagnosing FP using a mobile application with a user interface. The data used were 14 people with FP and 31 healthy people, and the accuracy of the diagnosis was 95.5%. In another study by Haase et al. [17], an analytical system was used to analyze the two sides of the face in detail, which is the coding system (FACS) using a dataset of 299 people with FP and 28 healthy people. However, in the training process, data from healthy subjects spoofed with facial paralysis were used. The analysis and scan duration was 108 ms per image. A study by Ngo et al. [18] proposed a method of limited-orientation modified circular Gabor filters (LO- MCGFs) to perform quantitative analysis of FP where a database of 85 subjects (75 patients + 10 healthy volunteers) was used to achieve 81.2% accuracy by utilizing a method based on frequency to preprocess images prior to extracting features. Wang et al. [19] proposed a technique to assess the level of FP, taking into account both the fixed facial unevenness and the variable transformative factors. A database with 62 patients, with 33 females and 29 males, using the suggested approach that integrates both static and dynamic quantification could achieve a recognition rate of 97.56%. Another study by Codari et al.

[20] proposed a method of facial thirds–based evaluation to measure the degree of facial asymmetry using a stereophotogrammetric device. The data used were from 40 healthy people and 30 people with FP. A hybrid method proposed by Storey et al. [21] created a three-dimensional image automatically from a two-dimensional image using a computer vision system based on 3D CNN to detect facial features, which achieved high accuracy in collecting facial features. However, there was an error that affected the accuracy of the system in collecting mouth features due to asynchronous movements. The study used two sets of data, and the F1 score was 88% for FP and reduced to 82% for mouth movement. Another study by Storey et al. [22] detected facial features and diagnosed FP by training a deep neural network using mixed training data of healthy and facial palsy subjects based on a binary sequential process proposed in [23], where the error classification ranged between 8.72% and 18.88%, and the accuracy of the evaluation ranged from 82% to 95.60%. Another study by Barbosa et al. [24] used training data consisting of 440 images of FP patients to extract features of FP by a set of regression trees for regularized logistic regression detection and iris detection with acceptable accuracy. Jiang et al. [25] used machine learning methods (K-NN, SVM, and NN) to classify the injury degree of FP by computational image analysis using a data set of 80 participants, with an average of 100 images for each person. The accuracy of the system ranged between 87.22% and 95.69%. Dell'Olio et al. [26] used the FaraPy system to detect FP in real time with different facial expressions for six healthy people, and with the lowest percentage of losses obtained an acceptable accuracy. To detect 68 facial features and diagnose FP, a complete CNN system by Liu et al. [27] was trained using a dataset of healthy subjects with different facial expressions and subjects with FP, and it extracted the facial features for classification. However, a major limitation was the length of execution time. A study by Nguyen et al. [28] detected different facial expressions based on three dimensional point cloud data and deep engineering learning technology to obtain a detection accuracy from 69.01% to 85.85%. Recently, Dominguez et al. [29] used 480 images of data, facial features, and their binary classifier to detect FP disease, where classification accuracy ranged between 94.06% and 97.22%. Another study by Estomba et al. [7] worked on predicting facial nerve palsy through the K-nearest neighbor algorithm using 356 patients' data, and the accuracy of the system performance exceeded 0.9. A study by Amsalam et al. [3] using a technique for detecting FP using a computer vision system was introduced. This method utilized deep learning through CNN and a Python program and involved the analysis of 570 images, including 200 images of individuals with facial palsy. The study included 10 participants, comprising three males and seven females with varying degrees of FP and injury on different sides, aged between 15 and 70 years. The method demonstrated short processing and detection time with 98% accuracy. However, it was not considered to be

user- friendly due to its lack of real-time functionality. In another study by Zhang et al. [30] a system that automatically evaluates faces called AFES (Automatic Facial Evaluation System) was suggested with customization options; during the study, a total of 92 individuals with facial palsy were enrolled and underwent evaluations. The evaluations included both subjective manual assessments using scales such as mHBGS and mSFGS, as well as automatic objective evaluations utilizing AFES. The AFES evaluations included aHBGS, aSFGS, and assessments of specific facial features. These evaluations were conducted at the beginning of the study and repeated after two weeks. AFES's algorithm was developed by training and testing it on video frames from over 100 patients; 80% of the frames were used as training sets, while the remaining 20% were used as testing sets. AFES may be viewed as a feasible approach for conducting a precise and dependable assessment of patients who have facial palsy. However, previous methods for diagnosing FP or detecting facial features had problems, including considering older methods [15,16] and often using training data from spoofed FP [23], Or the dataset of patients with FP may be small [26]. In addition to insufficient or non-existent quantitative results and low accuracy in many cases [17,31]. Therefore, this study proposes a high-accuracy diagnostic system to detect FP and the patient's gender and age in real-time by using deep learning algorithms based on CNN.

The remainder of this paper is as follows: Section 2 describes the materials and methods of the study, including research ethics and participants, experimental setup, system design, and Features Extraction. Section 3 includes the performance metrics of the proposed system. Section 4 presents the results and analysis of the proposed system performed on human participants with discussion. Finally, Section 5 concludes the paper

II. MATERIALS AND METHODS

In this study, the guidelines and research ethics issued by the Declaration of Helsinki in 1964 in Finland were followed. Approval from the specialized research committee in the Department of Research and Knowledge in Dhi Qar Health of the Iraqi Ministry of Health was obtained to conduct the research, according to protocol No. (363/2022). Prior to conducting the research, the consent of all participants was obtained to collect samples, and they were informed of the protection of their data.

The number of participants was 20 male and female individuals with different degrees of injuries. With a difference in the side of the injury, the average age ranges from 10 to 65 years. Samples were collected from the Department of Physiotherapy at Al-Rifai General Hospital. The remaining data were obtained from the Kaggle website [32], including UTKFace dataset [33] for normal people and FER-2013 dataset [34] for facial palsy. The dataset collected was 20,600 images consisting of 19,000 normal images and 1600 palsy images.

2.1. Experimental Setup

The experimental setup is capable of extracting features from the face and diagnosing FP, as shown in Figure

1. The patient sits in front of the camera at a distance of 0.5 to 2 m while the microcomputer (a Raspberry Pi) works to detect facial palsy and diagnose the condition in real-time, indicating whether the person is healthy or has FP, and determining the affected side of the face, as well as revealing the person's gender and age. The process of detecting facial features and diagnosing FP is conducted after training the system with three types of data, including right palsy data, left palsy data, and data from healthy individuals. The data in each case was divided into two parts: 80% training data and 20% test data, comprising 19,000 normal images and 1600 images of FP. The diagnostic process was carried out using the Python programming language (Version 3.9), and the code was executed on Anaconda (version 2.3.2) on the computer after installing image libraries and object detection libraries (OpenCV, dlib, TensorFlow, Keras, and NumP).
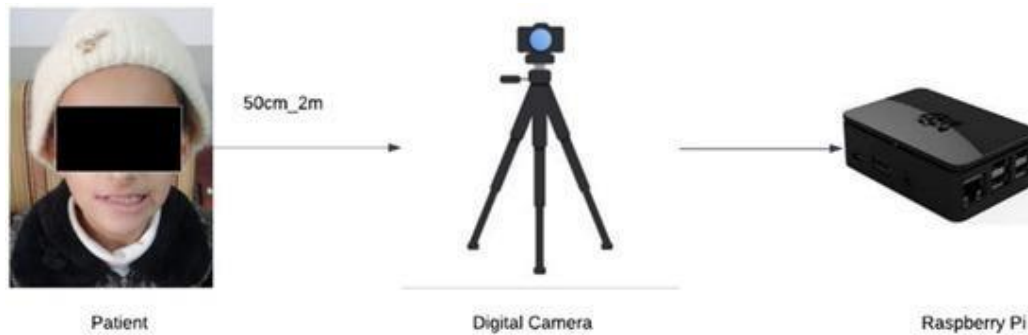
Figure 1. The real-time proposed system.

2.2 Hardware

The proposed system's practical part is shown in **Figure 2**, which comprises a Raspberry Pi, a digital camera, and a display screen mounted on a tripod.

Figure 2. The practical part of the proposed system.

Figure 2 shows the whole proposed system where the patient sits in front of the camera, the captured image is displayed on the screen, and the detection window appears around the face to complete the diagnosis process.

The Raspberry Pi 4 Model B 2G is a small single-board computer developed by the Raspberry Pi Foundation. The dimensions of the Raspberry Pi 4 Model B 2G are 88 × 58 × 19.5 mm (3.46 × 2.28 × 0.77 inches), which makes it small and portable. This makes it easy to use in projects and applications that require a small footprint, such as home automation, media centers, and portable gaming devices. It is important to note that these dimensions are for the board itself and do not include any additional components or peripherals that may be required for a specific application.

Raspberry Pi 4 Model B 2G is not a standalone device; it needs a power supply, SD card, keyboard, mouse, display, and other accessories to function effectively.

The display was a Wave Shares 5in HDMI Display Capacitive Touch Screen that with resolution of 800 × 480 pixels. The physical size of the screen is measured diagonally from corner to corner. The aspect ratio of the screen is 800 × 480, or approximately 1.67:1.

The camera used was a Kisonli (NO: U-227), with digital zoom (f = 3.85 mm) and 10 megapixels. After the code was installed on the Raspberry Pi device, the device was connected to a digital camera to photograph the patient and a screen to display the patient's image, and through it, the Raspberry Pi device was controlled. However, some additional considerations were attended to, such as directing the patient towards the camera and ensuring good lighting in the place of imaging to facilitate the process of detection and diagnosis of FP by the proposed system.

2.3 System Design

The block diagram of the proposed diagnostic system based on CNN is shown in **Figure 3**.
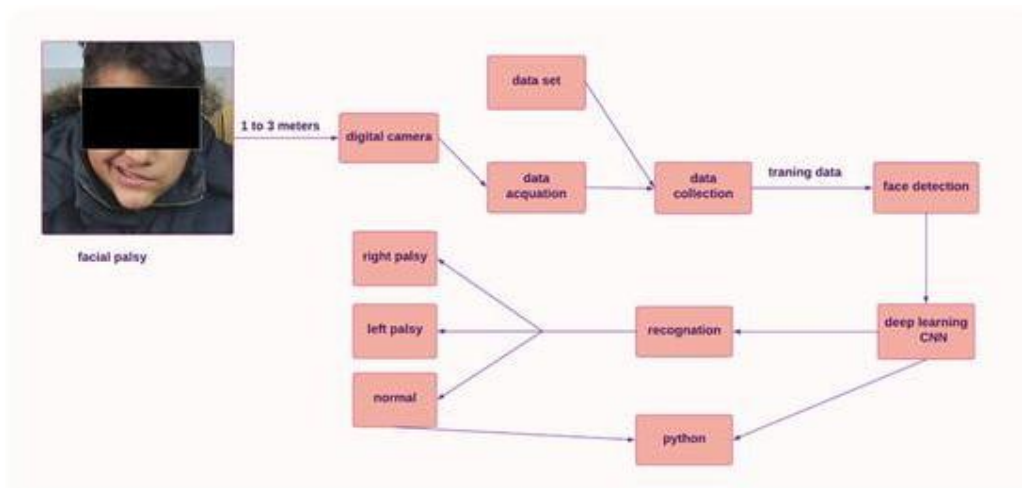

Figure 3. The block diagram of the proposed diagnostic system design.

2.4 Features Extraction

After the collected data is saved to a computer file, the FP classification process begins. The process of classifying facial features goes through three successive stages: face detection, extraction of facial features, and classification of expressions. The image was processed before detecting the face and extracting features, as it worked to detect the main features of the face, such as eyebrows, eyes, and mouth [35]. The main goal of feature extraction is to obtain the most relevant information from the original data and represent that information in a reduced dimensional space [36]. The proposed diagnostic system extracted 68 facial features using Haar Cascades, which is an popular technique used to quickly detect objects, including human faces [37]. Facial features were extracted in the proposed diagnostic system using rectangles to cover each facial feature. The rectangle was divided into two halves, one white and one black, due to differences in contrast for each area. The two regions were collected, and the result was extracted by the subtraction method. The closer the result of the subtraction is to 255 or one, the stronger the indicator of a particular trait. To reduce the processing time, the image was converted into an integrated image, and a detection process was applied to it. The capacity and size of the window were adjusted depending on the size of the feature, and the percentage likelihood of a person's presence increased with the increase in Haar's features. Facial features, such as eyes, eyebrows, facial edges, mouth, and nose, were detected using a set of packages in a Python program. Haar Cascade searches for similarities between images to match and detect differences between them, which increases detection accuracy. Classification is based on the presence or absence of these features and is classified as true or false.

2.5 Evaluation Metrics

Diagnostic results can be divided into 4 cases depending on the combination of actual and expected categories, namely: true negative (*TN*), true positive (*TP*), false negative (*FN*), and false positive (*FP*). Depending on the confusion matrix, the diagnostic ability of the proposed diagnostic system can be evaluated through five variables: sensitivity, specificity, precision, accuracy, Matthews' correlation coefficient, and error rate.

These variables can be defined as follows [38,39,40,41]:

$$Sensitivity = TP/(TP+FN) \tag{1}$$

$$Specificity = TN/(TN+FP) \tag{2}$$

$$Precision = TP/(TP+FP) \tag{3}$$

$$Accuracy = (TP+TN)/(TP+TN+FP+FN) \tag{4}$$

$$MCC = (TP \times TN - FP \times FN)/\sqrt{((TP+FN)(TN+FP)(TP+FP)(TN+FN))} \tag{5}$$

$$Errorrate = (FP+FN)/(TP+FP+FN+TN) \tag{6}$$

The confusion matrix is a graph that provides a complete visualization of the performance of a supervised deep-learning algorithm. Each row in the matrix represents the actual class states, and each column represents the expected class states. Errors can be easily calculated and visualized by determining the values that appear outside the vertical diagonal of the matrix, while valid values can be computed from the values that appear on the diagonals of the table [42]. The confusion matrix has been used in machine learning to explain and evaluate models' behavior for the supervised classification [43].

Case Studies/Experiments:

Case Study 1: E-commerce Platform

In the first case study, we tested an e-commerce platform dealing with a huge product catalog and patron statistics. Using MongoDB, we applied indexing on product categories and client IDs, optimizing the retrieval of product pointers for man or woman customers[5].

Case Study 2: Healthcare Database

In the second one case observe, we focused on a healthcare database containing affected person statistics and scientific histories. We leveraged MongoDB's aggregation framework to analyze affected person statistics for clinical research. This allowed us to aggregate and clear out records effectively to become aware of patterns and correlations.

Our experiments verified that the aggregation framework reduced the time required for complicated facts evaluation, making it simpler for researchers to get right of entry to and derive insights from the database.

Case Study 3: Content Management System

Our 0.33 case study revolved round a content control gadget dealing with various content types, along with articles, snap shots, and consumer-generated content. Here, we explored the benefits of schema design through a number of the information to streamline statistics retrieval. By doing so, we discovered a first-rate discount in question complexity, main to faster content material retrieval and progressed machine responsiveness.

III. EXPERIMENTAL RESULTS AND DISCUSSION

To ensure that the proposed diagnostic system achieved the desired goal and the validity of the results, we compared it with the data of 20 patients diagnosed by a physician. After training the system on the training data, the system's diagnostic accuracy was 98% and 99% for the case study data. The training set accounted for 80% of the data, while 20% was reserved for testing. The training time depends on the size of the data, The more data, the longer the training time, and vice versa. It also depends on the power of the computer processor used. For the proposed diagnostic system data, the training period was only five hours for one time, and the diagnostic time was only a few seconds. The accuracy of the proposed diagnostic system can be clearly seen in **Figure 4** and **Figure 5**, which show the results of the diagnosis. The system classified the condition of the person and detected the affected side of the face if the person had FP or not
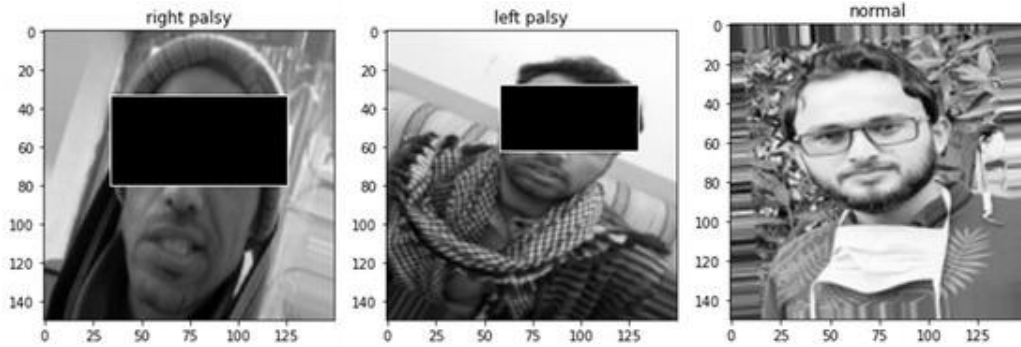
Figure 4. Result of the proposed diagnostic system for different male cases.
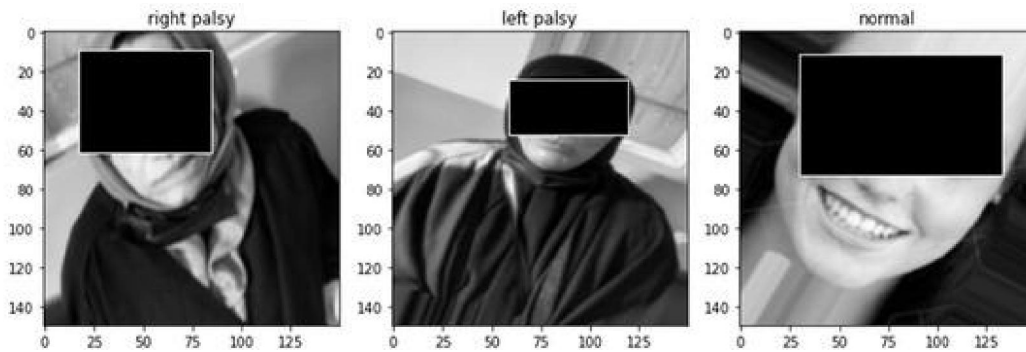

Figure 5. Result of the proposed diagnostic system for different female cases.

When the proposed system was applied for real-time diagnosis using the Raspberry Pi, it displayed the person's condition, the afflicted side of the patient, their gender, and age, as shown in **Figure 6**.
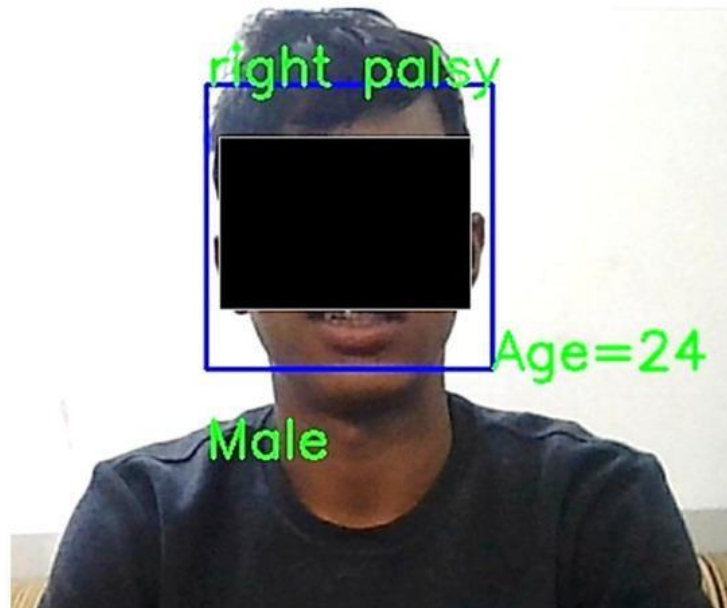

Figure 6. Real-time right FP patient diagnosis by the proposed system with age and gender detection.

The proposed system accurately diagnoses FP, identifies the affected side of the face (which was the right side), and can determine the patient's gender (in this case, male, approximately 24 years old).

The proposed system can also diagnose the condition of a healthy person and predict their gender and age, as shown in **Figure 7**.
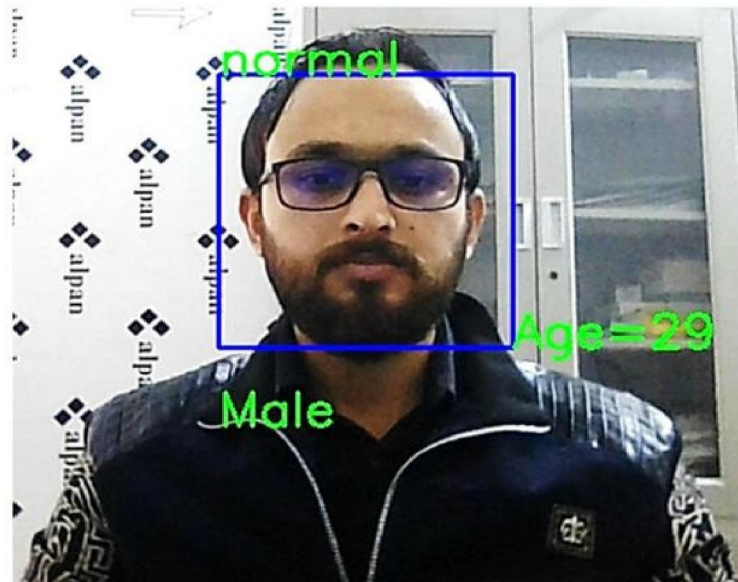

Figure 7. Real-time normal person diagnosis by the proposed system with age and gender detection.

The proposed system can also diagnose multiple people simultaneously and detect their gender and age, as shown in **Figure 8**

Figure 8. Diagnosis process for two people.

All the raw information for the prediction results of the model used in the dataset is in the confusion matrix. **Figure 9** represents a triangular matrix where the rows represent the actual values, and the columns represent the predicted values of the model. It shows the true positive, false positive, true negative, and false negative values and their compatibility with the actual values in the stored data

Figure 9. Classifications of the confusion matrix.

The confusion matrix is based on the classification of the test data, where the number of rows and columns in the matrix is equal to the size of the test data. **Figure 10** shows the confusion matrix in detail, which was obtained after training the data for 100 epochs. The test data was divided into three categories: right palsy, left palsy, and normal. Therefore, the confusion matrix appeared as a triangular matrix with three columns and three rows. The error data represents the number of instances that were not recognized by the program, while the accuracy data indicates the instances that were recognized by the program. The true and predicted values were identified by the confusion matrix. The accuracy and effectiveness of the system increased as the prediction accuracy increased. When the system's prediction results matched the stored data and their classification, the model could be adopted. The results indicated the acceptability and efficiency of the system in diagnosis, achieving an accuracy rate of 0.98 with a 0.02 error rate. The sensitivity, specificity, and precision of the system were 1, 0.8, and 0.97, respectively
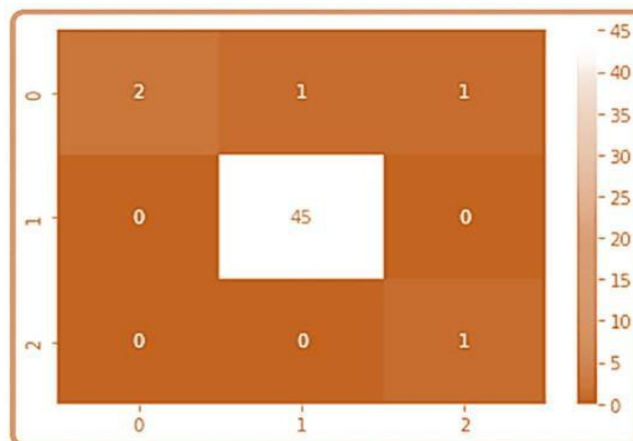

Figure 10. The confusion matrix.

250

To increase accuracy and reduce system losses, it is necessary to standardize the size of images, expand them, and increase the amount of data used. The high accuracy of the proposed system, which is 98%, indicates that it is an acceptable diagnostic system with few errors. Typical results of the proposed system are shown in **Figure 11**.
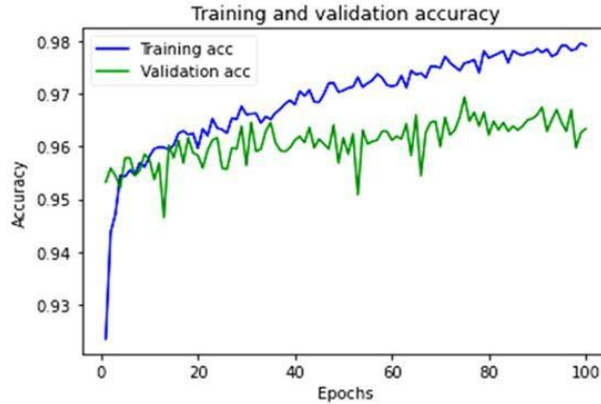

Figure 11. Training and validation accuracy

System losses decrease with increasing accuracy because the relationship between them is inverse. The percentage of losses in the proposed system was 2% of the total data used in the training process. **Figure 12** shows the percentage losses for training and validation in CNN training.
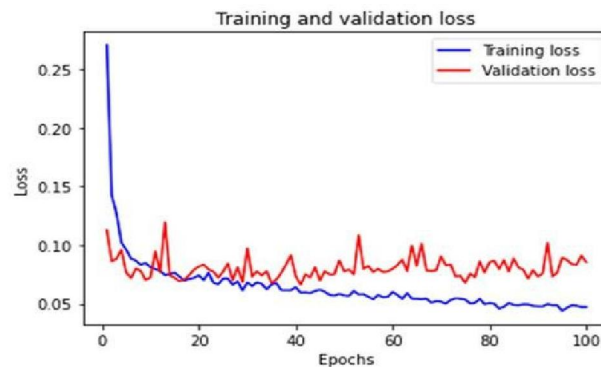

Figure 12. Training and validation loss.

The distinction between the prior techniques and the proposed system's approach can be clarified by examining the technology employed, the quantity of data utilized, and the system's precision, as illustrated in **Table 1** below.

Table 1. Shows the comparison between previous studies and the proposed system.

The proposed system outperforms previous studies in many features, including large numbers of training data, real-time diagnosis, and high accuracy, in addition to real-time gender and age detection and use of a practical Raspberry Pi device.

Despite the desirable features and accuracy of the proposed system in diagnosing FP and detecting gender and age, it has some limitations in diagnosis. These include the variation in the shapes of people of the same age in relation to determining the age, the difficulty in diagnosing when the patient moves, and the difficulty in distinguishing if the person suffers from a facial deviation resulting from an accident or a deviation in the nose. Furthermore, there are other limitations, such as the challenge of collecting data for disease cases because patients do not want to be imaged because of the embarrassment caused by FP.

V. CONCLUSION

In this paper, a modern and highly accurate diagnostic system to automatically detect facial paralysis that affects the seventh nerve in the face was proposed by us. The proposed system is based on CNN and can diagnose FP with high accuracy, along with detecting the patient's gender and age. The diagnostic accuracy of the proposed system reached 98%. It is suggested as an auxiliary medical diagnostic tool for doctors, nursing staff, and patients. The patient's use of this system at home in the diagnostic process reduces embarrassment, effort, time, and cost. Further work is ongoing to develop the system to diagnose more conditions.

REFERENCES

[1]. Amsalam, A.S.; Al-Naji, A.; Daeef, A.Y.; Chahl, J. Computer Vision System for Facial Palsy Detection. *J. Tech.* 2023, *5*, 44–51.

[2]. Kumar, G.; Bhatia, P.K. A detailed review of feature extraction in image processing systems. In Proceedings of the 2014 Fourth International Conference on Advanced Computing & Communication Technologies, Rohtak, India, 8–9 February 2014; pp. 5–12.

[3]. Saxena, K.; Khan, Z.; Singh, S. Diagnosis of diabetes mellitus using k nearest neighbor algorithm. *Int. J. Comput. Sci. Trends Technol. (IJCST)* 2014, *2*, 36–43.

[4]. Caelen, O. A Bayesian interpretation of the confusion matrix. *Ann. Math. Artif. Intell.* 2017, *81*, 429–450.

[5]. Ahmed, A. When is facial paralysis Bell palsy? Current diagnosis and treatment. *Cleve Clin. J. Med.* 2005, *72*, 398–401